Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1281 – 1288

Procedia Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Knowledge Structuring and Reuse System Using RDF for Supporting Scenario Generation

Teruaki Hayashi*, Yukio Ohsawa

Department of Systems Innovation, School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, 113-8656

Abstract

A scenario is a series of information derived from data or related knowledge. Action Planning is a workshop method for creating scenarios, which participants resolve conflicts and reduce risks of taking actions, through the process of externalizing and serializing knowledge. In this study, we show our work to create a database for reuse and retrieval of knowledge by converting scenarios created in Action Planning into structured knowledge with RDF. We store information of scenarios, and design the interface for helping users extract accurate information and make decision. By not only structuring knowledge used in the past Action Planning, but also making it possible to access external collective intelligence of DBpedia and Data Jacket store, this system is expected to retrieve important information about solving problems.

Peer-review under responsibility of KES International

Keywords: market of data; scenario; RDF; Innovators Marketplace; Data Jacket

* Corresponding author. Tel.:+81-5841-2908.
*E-mail address:*teru-h.884@nifty.com

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2015.08.233

1. Introduction

1.1. Expectations and problems for data utilization

The potential expectation has been increased about generating innovative businesses or creating values by combining existent data from different regions across organizations and sectors. By the rapid spread of personal devices, e.g., smart phones, and Social Networking Services, e.g., twitter or facebook, the amount of personal data such as purchase histories or moving records comes to be able to be stored, which had been said difficult to be acquired. In addition to the conventional data acquired in the industries, it becomes possible to acquire highly precise data because of the development of sensors. Some companies and organizations have taken advantage of the data in-house, and have started to improve their existent services by adding value. Furthermore, by combining the data from different regions across organizations and sectors, and acquiring new knowledge, a potential possibility to generate innovative business has been increased. On the other hand, national or local governments have been supplied their datasets to the public allowing the secondary use on the Web for encouraging the reuse of data (Open Data). Linked Open Data (LOD) has been a recent movement to interlink a dataset with other open datasets[1]. LOD is based on a concept of Linked Data, i.e., publishing and connecting structured data on the Web, originally proposed by Berners-Lee[2]. Although LOD has attracted social attention and expectation for data utilization, datasets in private companies, most autonomies or individuals are not open or shared due to the cost of data management and security issues. People hardly publish their personal data because of the issue of personal identifiable information and information security, e.g., daily schedule or medical inspection data. Companies hardly open in public their data for fear of loss of business opportunities. Therefore, companies or individuals may hesitate to share data to the general public.

1.2. Necessity for creating a market of data

Although there are expectations for data utilization among private sectors or organizations, various problems are pointed out as mentioned above. Accordingly, instead of forcing to open or share data, it is important to create a market of data. The market of data is not about data mining for marketing, but the platform for innovations based on data. In the market of data, users select data which they want, negotiate with data owners, and get data at a reasonable conditions, e.g., price for exchanging datasets, based on the principle of free market. The market of data is a platform for innovations among stakeholders, not only deployed on the Web. The data portal platforms or services[3,4,5] to promote data exchange or buying/selling have been provided. However, it cannot be expected communication among data owners and users, only to list up the superficial information of data on the Web. The market of data should be a platform for negotiating, considering, and obtaining valuable datasets.

2. Process for activating a market of data: Innovators Marketplace on Data Jackets (IMDJ)

In order to create a Market of Data, it is necessary to establish the platform for sharing the information about data, i.e., meta-data. Innovators Marketplace on Data Jackets[6,7] (hereafter IMDJ) is a gamified workshop for discussing the data utilization, by providing information about data as Data Jackets. A Data Jacket[8,9] (hereafter DJ) is a summary of datasets, that is, meta-data. We can understand the outline, format, or variables of data referring to metadata on DJs, even if data itself is not open. By publishing DJs, participants of IMDJ can consider the contents of datasets by reading DJs, and start to communicate about data utilizations among data owners, analysts and data users. Although private companies have collected and stored amounts of datasets, there is no basis to share or open data considering management cost and security, and most of their data are closed. DJ is a technique for enabling to share information about data and consider potential usefulness of datasets, allowing keeping data itself hidden. DJs have been currently registered more than 510 (April, 2015).

In the process of IMDJ, we introduce tools for data visualization, e.g., KeyGraph[10], and create a map on which shows possible combinations among DJs, which support participants to discover latent combinations of each dataset. Data owners provide their datasets as DJs, and participants of IMDJ (including data owners, users, and analysts) create solutions for solving data users' problems which are stated as requirements. Through the communication among participants, data owners are able to understand how to use their own data from possible combination of DJs proposed by data analysts, and users will understand how their requirements can be satisfied with proposals (we call "solutions" in IMDJ). Through the process of IMDJ, participants start to negotiate for data exchange or buying/selling to create new businesses.

As the post-process of IMDJ, Action Planning[11,12] for refining solutions into practical scenarios of actions has been proposed. Through the process of Action Planning (hereafter AP), solutions created in IMDJ are grounded into feasible plans, by adding related data, elements, knowledge, or stakeholders into scenarios, which participants resolve conflicts and reduce risks of taking actions, through the process of externalizing and serializing knowledge.

Currently, Data Driven Innovation Strategy Council organized by Ministry of Economy, Trade and Industry of Japan introduced our methods of IMDJ and AP for creating businesses by involving 54 enterprises. In other project, Data Exchange Consortium, which is a coalition of private companies, has been utilized IMDJ and AP for exchanging the information of datasets among stakeholders and standardization of the data collaboration support system.

However, scenarios created in AP are only discussed or shared among participants of workshops. Although entered DJs and use cases (solutions and requirements) are published online on Data Jacket Store (explained later), the detailed information about scenarios, e.g., stakeholders or resources included in scenarios, is not shared with general public. In order to activate a market of data and data-driven innovations, it is considered important to make available the knowledge included in scenarios, which have been created in the past discussions.

In this paper, we show our work to create a database for reuse and retrieval of knowledge by converting scenarios created in AP into structured knowledge. We create a knowledge base by describing stakeholders and resources used in the past AP workshops with RDF, which is a framework of language for metadata description. We store information of scenarios (stakeholders and resources), and design the interface as a system which recommends relevant stakeholders and resources, when users create new scenarios. Moreover, utilizing the characteristics of the RDF, by connecting proposed system with Data Jacket Store and DBpedia as an external database, it is possible to recommend relevant knowledge to user's scenario.

3. Structuring Knowledge Constituting Scenarios

3.1. Outline of Action Planning

A scenario is a series of information derived from data or related knowledge. Human decision makers read, interpret, and perform actions according to the scenarios. AP is a workshop method for creating scenarios, which is designed for formulating a discussion and leading atypical viewpoints and knowledge. Referring to knowledge, participants resolve conflicts and reduce risks of taking actions, through the process of externalizing and serializing knowledge. AP has the following three phases for refining solutions proposed in IMDJ into scenarios.

(1) Requirement Analysis phase (AP1): Requirement Analysis means the phase of acquiring covert requirements which targets do not recognize yet, through the discussion. Starting from targets' overt requirements, participants find the background factors of overt requirements from objective data or suppositions with logical thinking, and clarify the covert requirements and potential stakeholders.

(2) Element Externalization phase (AP2): Externalizing related knowledge or information for creating solutions from the requirements or solutions clarified in AP1. Externalized knowledge includes cost, time to realize solutions, resources (technologies, equipment, budget), stakeholders (targets, supporters, dissidents) and shared advantage or countermeasure.

(3) Element Serialization phase (AP3): Serializing the knowledge and information considering relations to derive the satisfaction of the requirements or the solution of the issues. Serialization means to find relationships among knowledge/information, and connect them to form a scenario by following a particular rule sets (time management, business modeling, etc.).

The scenario generation of AP proceeds by filling sheets (one example of the sheets is shown in Fig.1). By planning actions resolving conflicts among participants with various viewpoints, created scenarios help participants perform and reduce risks of taking actions, i.e., creating businesses.

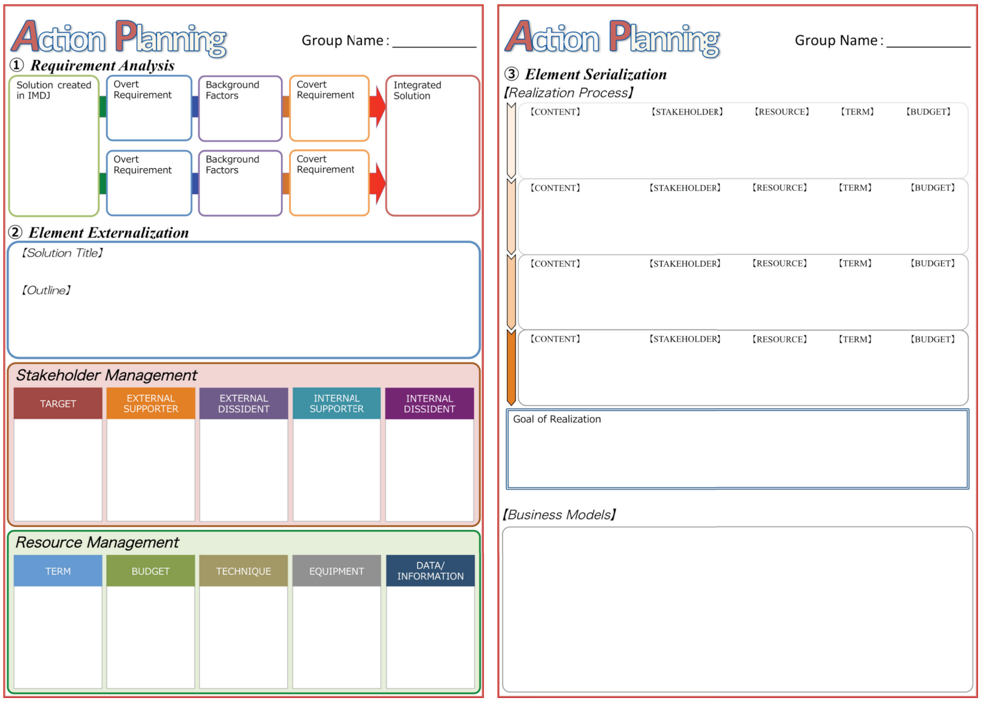

Fig. 1. Example of Action Planning Sheets.

3.2. Connecting knowledge by RDF

RDF (Resource Description Framework) is a labeled directed graph data format, and a framework of language for metadata description, which is a core technology of the Semantic Web. In RDF, the combination of subject, predicate, and object constitutes the basic description unit, which is called RDF triple. RDF triple constitutes a single sentence. By combining RDF triples, the network structure of the knowledge, concepts, or the relationship between resources can be described.

For example, let us think about the sentence "The name of scenario:0001 is safe and secure route recommender system." To describe this sentence in RDF, The subject is "scenario:0001", predicate is "name", and object is "safe and secure route recommender system". In this case, the predicate "name" connects subject "scenario:0001" and object "safe and secure route recommender system" (Fig. 2). RDF description has labels not only on nodes, but also on edges. By connecting RDF triples, it is possible to describe complex knowledge represented as sentences.

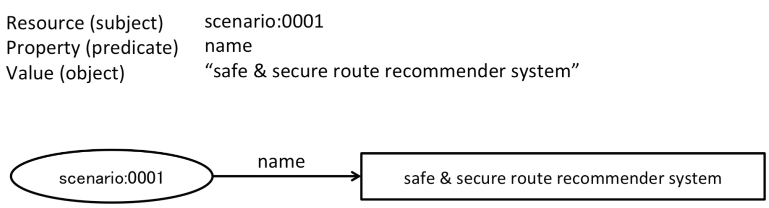

Fig. 2. One example of RDF description.

As shown in the example above, we structurally describe knowledge contained in scenarios in RDF description. Fig. 3 is the example of the scenario created in AP (extracting a part of stakeholders and resources), and Fig. 4 shows the RDF graph description of the scenario in Fig. 3. In this study, in order to make it simple to understand the structure of knowledge, we focus on the knowledge of stakeholders and resources associated with scenarios, which is introduced in AP2. A description of the semantic structure of the requirement analysis (AP1) and elements serialization (AP3) is currently developing, and we do not mention in this paper. In the rest part of this paper, we use the term "a scenario" as a construct of a name, outline, stakeholders, and resources.

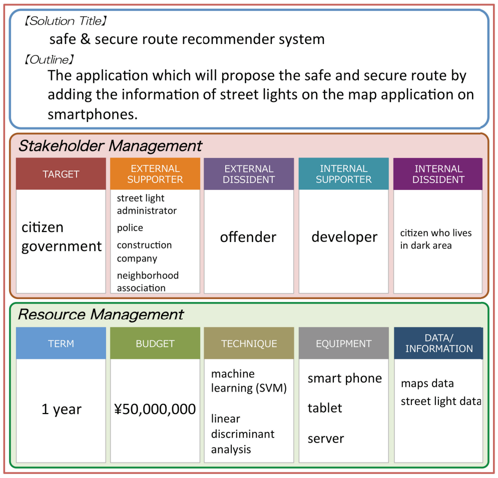

Fig. 3. A scenario created in Action Planning.

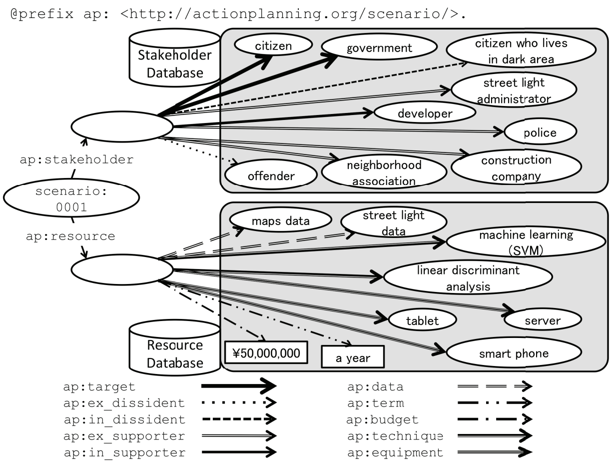

Fig. 4. RDF graph of a scenario shown in Fig. 3 (extracting a part of stakeholders and resources).

In this example, "ap:target" is a predicate representing target in the scenario. For example, a scenario has stakeholders "citizen" and "government" as targets (Fig. 3). When we represent it as sentence, it may be written as "Scenario:0001 has a stakeholder, and the targets are citizen and government." This sentence can be described as RDF graph using a blank node (Fig. 4), which can represent RDF graph with multiple values. The advantages of introducing RDF for describing the components of scenarios are that RDF description enables to represent the relationships between a scenario and the values constituting scenario with a simple rule; subject, predicate, and object. Moreover, in case a scenario has multiple targets, such as “citizen” or “government,” even if a new target “child” is added to the scenario, it is only to add a predicate “ap:target” and an object “child” from the blank node connected with predicate “ap:stakeholder”. The description of RDF graph format compensates for the shortcomings of table format, because in case of an addition of elements to scenarios, it can be solved only by adding nodes and links as an extension of structure, instead of adding columns for additional elements in table format.

4. Connecting Elements in Scenarios with other Databases

4.1. Connecting with Data Jacket Store and DBpedia

In the last chapter, we discussed the advantages of RDF description in structuring stakeholders and resources in scenarios. In this chapter, we discuss knowledge structuring by linking with other databases, which makes it possible to search and discover related knowledge with scenarios.

Because RDF describes the structure of data with nodes and links, it has the feature that can have links to resources in other databases. We combine the RDF store of scenarios with Data Jacket Store[13,14] and DB pedia[15]. DJs are currently described in RDF. Data Jacket Store (hereafter, DJ Store) is an information retrieval and recommender system for Data Jackets, which supports to search accurate information of datasets and their use cases. Not only reusing and structuring information contained in DJs, but reusing and structuring solutions and requirements stated in past IMDJ workshops, it is expected to be possible to retrieve important information about solving problems, which users could not notice by themselves.

We reuse URIs (Uniform Resource Identifier), which are already used in DJ Store, and RDF store of scenarios acquires links with external databases; DJ Store. For example, in the scenario of “safe and secure route recommender system”, the needed datasets are “street light data” and “map data”, which are resources of the scenario. By using URIs already used in DJ Store, it is possible for resources in RDF store of scenarios to acquire links with resources in external databases. Assuming that URI of “street light data” in DJ Store is “http://datajacket.org/datajacket/0150”, and URI indicates “map data” to be “http://datajacket.org/datajacket/0342”. The graph shown in Fig. 5 means that the needed datasets of “scenario:0001” is “http://datajacket.org/datajacket/0150” and “http://datajacket.org/datajacket/0342”, and the name of the dataset (the predicate which means a name is “rdfs:label”) “http://datajacket.org/datajacket/0150” is “street light data”, and the name of the dataset “http://datajacket.org/datajacket/0342” is “map data”.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@prefix ap: <http://actionplanning.org/scenario/>
@prefix dj: <http://datajacket.org/datajacket/>.
```

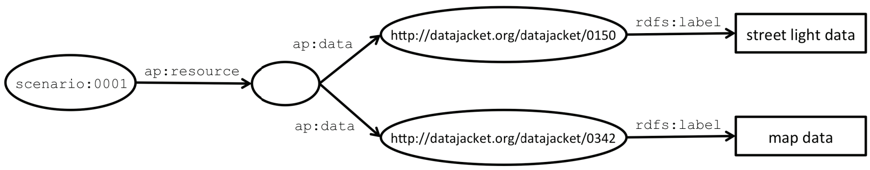

Fig. 5. Connecting RDF store of scenario with Data Jacket Store through URIs that is used in Data Jacket Store.

As with the same way of connection between proposed system and DJ Store, we introduce URIs utilized in describing resources in DBpedia. DBpedia is one of the applications of Linked Open Data, which connects multiple datasets on Wikipedia encyclopedia and enables users to discover further information using RDF links, which is the building block for the Semantic Web. DBpedia provides information on Wikipedia which are written in different languages. In this paper, we use DBpedia Japanese for connecting resources by RDF. In DBpedia Japanese, SPARQL Endpoint and the Web API are provided by Virtuoso server system. SPARQL (Simple Protocol And RDF Query Language) is a query language for RDF, which searches a set of triple patterns of RDF from RDF store.

For example, in the scenario of "safe and secure route recommender system", the stakeholders as targets are "government" and "citizen". The detailed information of these stakeholders has been stored in DBpedia. It is possible to access the applicable data by using URIs already used in DBpedia. In DBpedia Japanese, the resource representing "government" is "http://ja.dbpedia.org/resource/%e6%94%bf%e5%ba%9c", and "citizen" is "http://ja.dbpedia.org/resource/%e5%b8%82%e6%b0%91" (URIs are encoded in UTF-8). The graph shown on Fig. 6 represents that the stakeholders of "scenario:0001" as targets are "http://ja.dbpedia.org/resource/%e6%94%bf%e5%ba%9c" whose name is "government", and "http://ja.dbpedia.org/resource/%e5%b8%82%e6%b0%91" whose name is "citizen". Of course, not only the name of resources of URIs, but also a variety of objects are linked by predicates with URIs.

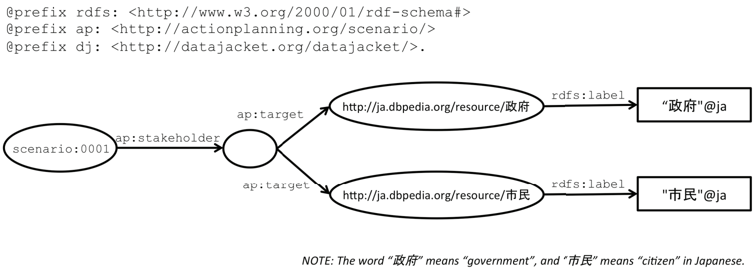

Fig. 6. Connecting RDF store of scenario with DBpedia through URIs that is used in DBpedia Japanese.

4.2. Recommendation of knowledge in scenario creation

Let us show an example of knowledge recommendation from the RDF store of scenarios. When a user inputs a new scenario on the Web browser, this system returns related knowledge retrieved from RDF stores. Fig. 7 shows relevant knowledge extracted from a new scenario which a user inputs. First, from the keywords contained in the input scenario, this system searches by sending SPARQL to the RDF store where the structured knowledge of the past scenarios is stored. When the related knowledge (stakeholders or resources) is found, the system returns them in a table format. "Category" represents that the recommended information is stakeholder or resource, and "type" shows the relationship between the knowledge and the input scenario. "Outline" is extracted from DJ Store and DBpedia, referring to external databases. A method for evaluating recommended knowledge is currently under development.

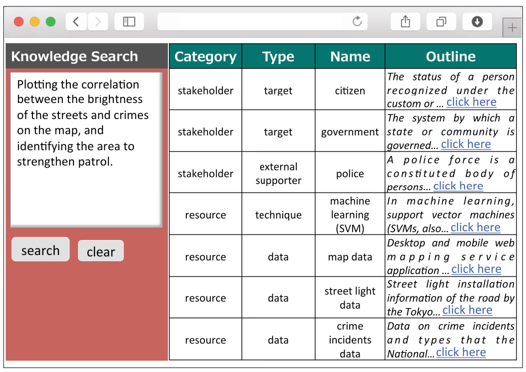

Fig. 7. Example of search results displayed on a Web browser.

5. Conclusion

In this study, we discussed a knowledge structuring and reusing system of scenarios created in Action Planning. We propose to create a database for accurate extraction and reuse of knowledge by converting scenarios with RDF. By introducing RDF, we structurally describe knowledge in scenarios. By storing structured knowledge, it is possible to implement a knowledge base for supporting users with creating scenarios in taking action by recommending knowledge. We build a system for access data and knowledge, not having been created in the past Action Planning, but also external collective intelligence of DBpedia and Data Jacket Store.

For example, in this system, when creating a new scenario, it is possible to recommend information about potential business partners and necessary data for users. In addition, the recommendation of knowledge, the stakeholders that may be dissident can be expected in advance. Thus, proposed system is considered to be able to help decision makers reduce risks of taking actions, and contribute to activating a market of data.

In this study, we implemented a system which extracting the resources or stakeholders from keywords included in the users' scenarios. However, if the keywords which have many links with related knowledge, knowledge which is recommended to users may be enormous. In the future work, form the point of view of accurate recommendation, we aim to build a system which will recommend the knowledge which is relevant to users in cooperation with the recommendation algorithm.

Acknowledgements

This study was partially supported by JST-CREST. Also we would like to thank to all the staff members of KKE (Kozo Keikaku Engineering Inc.) for supporting our research.

References

1. Yu L. *A Developer's Guide to the Smantic We*b. Springer; 2011.
2. Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific American; 2001.
3. Microsoft Azure Marketplace, [Online]. Available from: <https://datamarket.azure.com/browse/data>. [Accessed 20th April 2015].
4. CKAN, [Online]. Available from: <http://ckan.org/>. [Accessed 20th April 2015].
5. KDnuggets, [Online]. Available from: <http://www.kdnuggets.com/>. [Accessed 20th April 2015].
6. Liu C, Ohsawa Y, Suda Y. Valuation of Data through Use- Scenarios in Innovators' Marketplace on Data Jackets, ICDMW 2013, 1st Workshop on Market of Data, p.694-701.
7. Ohsawa Y, Kido H, Hayashi T, Liu C, Komoda K. Innovators Marketplace on Data Jackets, for Valuating, Sharing, and Synthesizing Data, Knowledge-based Information Systems in Practice, Springer-Verlag, 30, p.83-97, 2015.
8. Ohsawa Y, Kido H, Hayashi T, Liu C. Data Jackets for Synthesizing Values in the Market of Data. 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, Procedia Computer Science 2013; **22**, p.709-716.
9. Ohsawa Y, Liu C, Hayashi T, Kido H. Data Jackets for Externalizing Use Value of Hidden Datasets, 18th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, Procedia Computer Science 2014; **35**, p.946-953.
10. Ohsawa Y, Nels EB, Yachida M. KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor, Proc. Advanced Digital Library Conference (IEEE ADL'98); 1998, p.12-18.
11. Hayashi T, Ohsawa Y. Processing Combinatorial Thinking: Innovators Marketplace as Role- based Game Plus Action Planning, International Journal of Knowledge and Systems Science, **4**(3),2013. p.14- 38.
12. Hayashi T, Ohsawa Y. Estimatin of Novelty Assessment of Strategic Scenarios Using Relativeness, ICDMW 2014, 2nd Workshop on Market of Data. p.441-446.
13. Hayashi T, Ohsawa Y. Knowledge Structuring and Reuse System Design Using RDF for Creating a Market of Data, 2nd International Conference on Signal Processing and Integrated Networks; 2015. p.566-571.
14. Data Jacket Store, [Online]. Available from: <http://www.panda.sys.t.u-tokyo.ac.jp/hayashi/djs/djs4ddi/>. [Accessed 20th April 2015].
15. Auer S, Bizer C, Kobilarov G. DBpedia: a Nucleus for a Web of Open Data, In Aberer et al. (Eds.): The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference; 2007.
16. DBpedia Japanese, [Online]. Available from: <http://ja.dbpedia.org/>. [Accessed 20th April 2015].